

# MINI PROJECT

 **big**  
**basket**



## Exploratory Data Analysis

*Prepared by;*

**Data analytics**  
**Avinash kumar**  
avinash969658@gmail.com

# About the company

BigBasket is India's largest online grocery shopping platform, founded in 2011 in Bengaluru, Karnataka. The company was started by Hari Menon, V. S. Sudhakar, Vipul Parekh, V. S. Ramesh, and Abhinay Choudhari with the goal of making grocery shopping easy, fast, and convenient for people across India.

BigBasket allows customers to order fruits, vegetables, groceries, dairy products, household items, and personal care goods through its website and mobile app. It offers same-day delivery, next-day delivery, and quick delivery (BB Now) services in many cities.

In 2021, Tata Digital, a subsidiary of the Tata Group, acquired a majority stake in BigBasket. This helped the company grow rapidly and expand its services across 30+ cities in India.

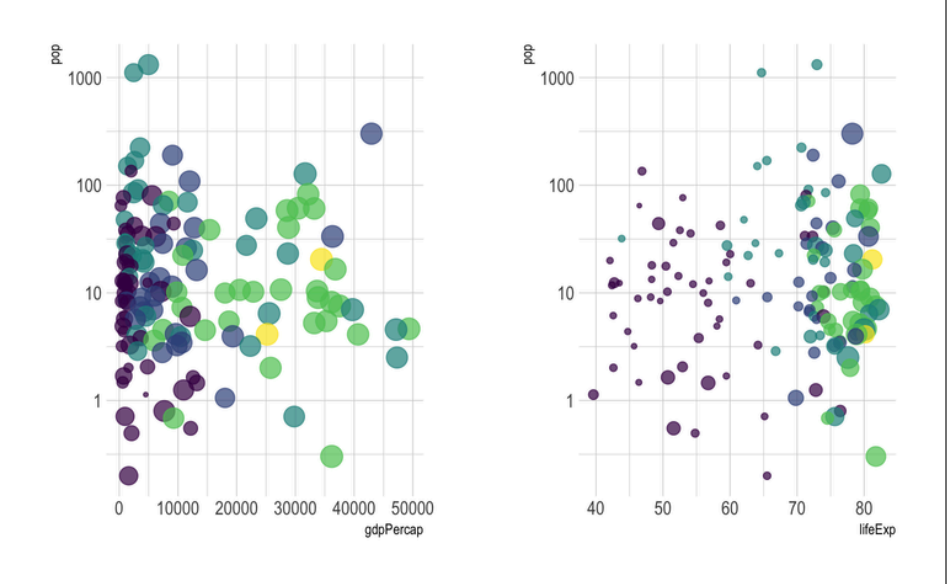
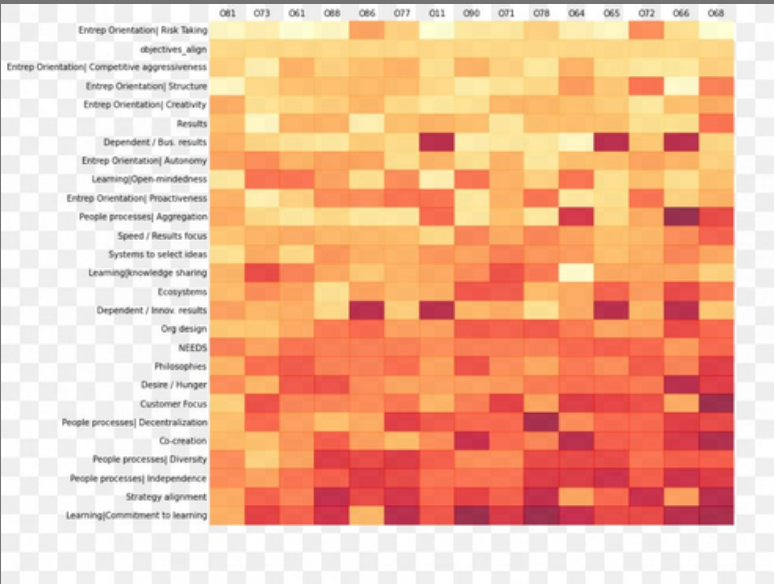
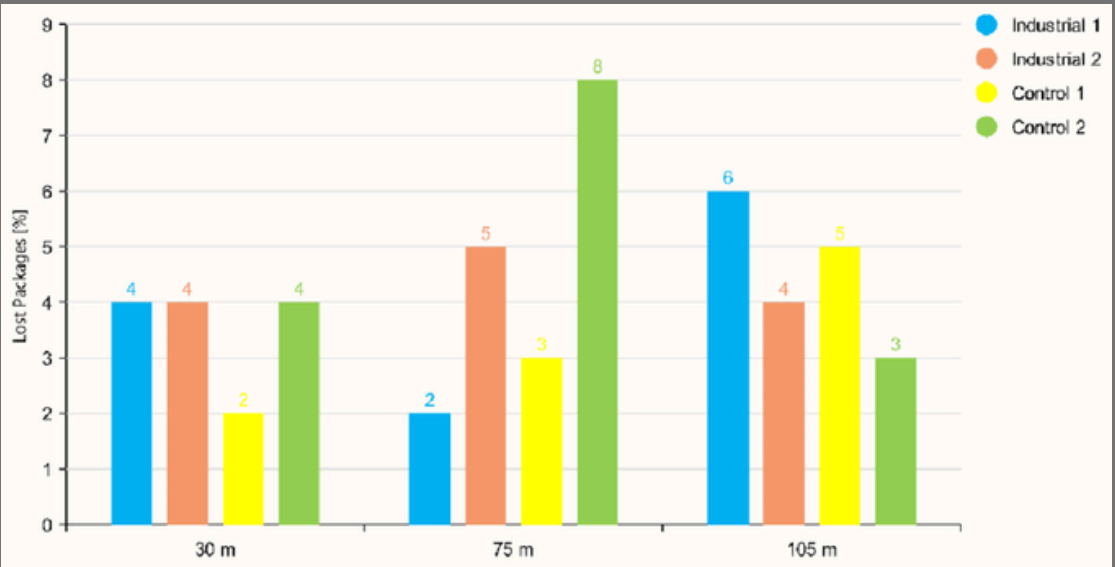
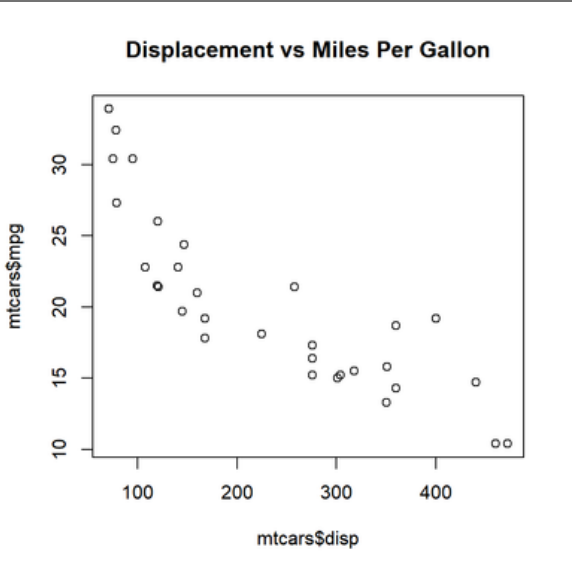
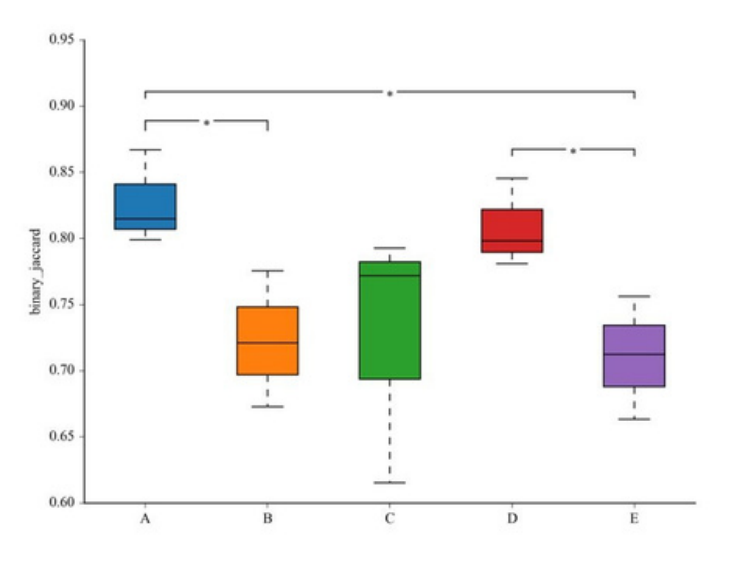
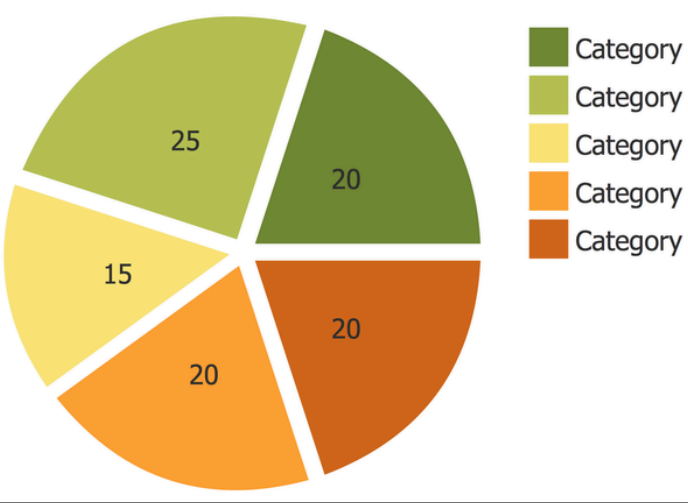
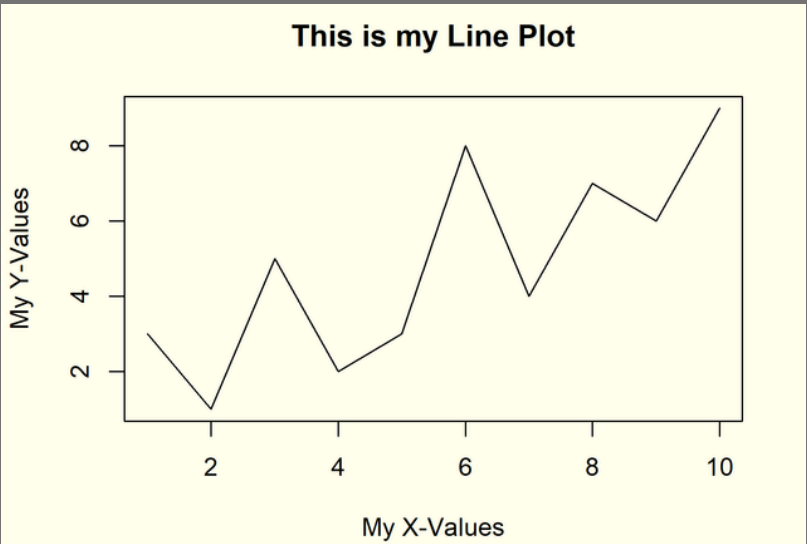
BigBasket also has its own brands like Fresho, BB Royal, and GoodDiet, offering good-quality products at affordable prices.

Today, BigBasket serves millions of customers, handles millions of orders every month, and continues to lead the online grocery industry in India through innovation, technology, and excellent customer service.

# Tools



# Graph




# Description

- **index** - Simply the Index!
- **product** - Title of the product (as they're listed)
- **category** - Category into which product has been classified
- **sub\_category** - Subcategory into which product has been kept
- **brand** - Brand of the product
- **sale\_price** - Price at which product is being sold on the site
- **market\_price** - Market price of the
- **product type** - Type into which product falls
- **rating** - Rating the product has got from its consumers
- **description** - Description of the dataset (in detail)

# Data reading

- Loading and Preparing **big basket** Dataset for Analysis in Google Colab

```
[ ] from google.colab import drive  
    drive.mount('/content/drive')
```

 Mounted at /content/drive

- in this analysis , we import **Pandas** for data manipulation ,**Numpy** for numerical operations ,**Matplotlib** for creating visualization ,and **Seaborn** and **plotly** for enhanced statistical graphic.

```
import numpy as np  
import pandas as pd  
import seaborn as sns  
import matplotlib.pyplot as plt  
import plotly.express as px
```

- We import the Google drive module to mount Google 'Drive', enabling access **to files and dataset in the for our analysis.**

```
df=pd.read_csv("/content/drive/MyDrive/Car vehicle/BigBasket Products.csv")
```

- We use pandas to Read the CSV file containing **big basket** Data from Google Drive, loading it into a DATAFRAME Name 'df' for analysis.

# Data overview

		index	product	category	sub_category	brand	sale_price	market_price	type	rating	description
0	1		Garlic Oil - Vegetarian Capsule 500 mg	Beauty & Hygiene	Hair Care	Sri Sri Ayurveda	220.00	220.0	Hair Oil & Serum	4.1	This Product contains Garlic Oil that is known...
1	2		Water Bottle - Orange	Kitchen, Garden & Pets	Storage & Accessories	Mastercook	180.00	180.0	Water & Fridge Bottles	2.3	Each product is microwave safe (without lid), ...
2	3		Brass Angle Deep - Plain, No.2	Cleaning & Household	Pooja Needs	Trm	119.00	250.0	Lamp & Lamp Oil	3.4	A perfect gift for all occasions, be it your m...
3	4		Cereal Flip Lid Container/Storage Jar - Assort...	Cleaning & Household	Bins & Bathroom Ware	Nakoda	149.00	176.0	Laundry, Storage Baskets	3.7	Multipurpose container with an attractive desi...
4	5		Creme Soft Soap - For Hands & Body	Beauty & Hygiene	Bath & Hand Wash	Nivea	162.00	162.0	Bathing Bars & Soaps	4.4	Nivea Creme Soft Soap gives your skin the best...
...	...		...	...	...	...	...	...	...	...	...
27550	27551		Wottagirl! Perfume Spray - Heaven, Classic	Beauty & Hygiene	Fragrances & Deos	Layerr	199.20	249.0	Perfume	3.9	Layerr brings you Wottagirl Classic fragrant b...
27551	27552		Rosemary	Gourmet & World Food	Cooking & Baking Needs	Puramate	67.50	75.0	Herbs, Seasonings & Rubs	4.0	Puramate rosemary is enough to transform a dis...
27552	27553		Peri-Peri Sweet Potato Chips	Gourmet & World Food	Snacks, Dry Fruits, Nuts	FabBox	200.00	200.0	Nachos & Chips	3.8	We have taken the richness of Sweet Potatoes (...)
27553	27554		Green Tea - Pure Original	Beverages	Tea	Telley	396.00	495.0	Tea Bags	4.2	Telley Green Tea with its refreshing pure, ori...

• Use head function to look for first 12 rows.

		index	product	category	sub_category	brand	sale_price	market_price	type	rating	description
0	1		Garlic Oil - Vegetarian Capsule 500 mg	Beauty & Hygiene	Hair Care	Sri Sri Ayurveda	220.0	220.0	Hair Oil & Serum	4.1	This Product contains Garlic Oil that is known...
1	2		Water Bottle - Orange	Kitchen, Garden & Pets	Storage & Accessories	Mastercook	180.0	180.0	Water & Fridge Bottles	2.3	Each product is microwave safe (without lid), ...
2	3		Brass Angle Deep - Plain, No.2	Cleaning & Household	Pooja Needs	Trm	119.0	250.0	Lamp & Lamp Oil	3.4	A perfect gift for all occasions, be it your m...
3	4		Cereal Flip Lid Container/Storage Jar - Assort...	Cleaning & Household	Bins & Bathroom Ware	Nakoda	149.0	176.0	Laundry, Storage Baskets	3.7	Multipurpose container with an attractive desi...
4	5		Creme Soft Soap - For Hands & Body	Beauty & Hygiene	Bath & Hand Wash	Nivea	162.0	162.0	Bathing Bars & Soaps	4.4	Nivea Creme Soft Soap gives your skin the best...
5	6		Germ - Removal Multipurpose Wipes	Cleaning & Household	All Purpose Cleaners	Nature Protect	169.0	199.0	Disinfectant Spray & Cleaners	3.3	Stay protected from contamination with Multipu...
6	7		Multani Mati	Beauty & Hygiene	Skin Care	Satinance	58.0	58.0	Face Care	3.6	Satinance multani matti is an excellent skin t...
7	8		Hand Sanitizer - 70% Alcohol Base	Beauty & Hygiene	Bath & Hand Wash	Bionova	250.0	250.0	Hand Wash & Sanitizers	4.0	70%Alcohol based is gentle of hand leaves skin...

• Get Description of the data in the DataFrame.

	index	sale_price	market_price	rating
count	27555.00000	27549.000000	27555.000000	18919.000000
mean	13778.00000	334.648391	382.056664	3.943295
std	7954.58767	1202.102113	581.730717	0.739217
min	1.00000	2.450000	3.000000	1.000000
25%	6889.50000	95.000000	100.000000	3.700000
50%	13778.00000	190.320000	220.000000	4.100000
75%	20666.50000	359.000000	425.000000	4.300000
max	27555.00000	112475.000000	12500.000000	5.000000

- Total records: 27,555
- Average sale price: ₹334.64
- Average market price: ₹382.06
- Average rating: 3.94 / 5
- Minimum sale price: ₹2.45 → Maximum sale price: ₹112,475.00
- Minimum market price: ₹3.00 → Maximum market price: ₹12,500.00

# Data information

- **Find Information about the DataFrame.**

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 27555 entries, 0 to 27554
Data columns (total 10 columns):
#   Column                Non-Null Count  Dtype  
---  -
0   index                 27555 non-null  int64   
1   product               27554 non-null  object  
2   category              27555 non-null  object  
3   sub_category          27555 non-null  object  
4   brand                 27554 non-null  object  
5   sale_price            27549 non-null  float64  
6   market_price          27555 non-null  float64  
7   type                  27555 non-null  object  
8   rating                18919 non-null  float64  
9   description           27440 non-null  object  
dtypes: float64(3), int64(1), object(6)
memory usage: 2.1+ MB
```

- **index 27,555 int64** Unique identifier for each product.
- **product 27,554 object** Name of the product (1 missing value).
- **category 27,555 object** Main product category (e.g., fruits, groceries, etc.).
- **sub\_category 27,555 object** Subcategory of the product (e.g., vegetables, snacks).
- **brand 27,554 object** Brand name (1 missing value).
- **sale\_price 27,549 float64** Selling price (after discount). 6 missing values.
- **market\_price 27,555 float64** Original price (before discount).
- **type 27,555 object** Type or classification of product.
- **rating 18,919 float64** Customer rating (many missing values).
- **description 27,440 object** Product description (115 missing values).

- \* **Measuring discount on a certain item.**

```
df["discount"] = df["market_price"] - df["sale_price"]
```

```
df.columns
```

```
Index(['index', 'product', 'category', 'sub_category', 'brand', 'sale_price',  
      'market_price', 'type', 'rating', 'description', 'products',  
      'discount'],  
      dtype='object')
```

```
df["discount"] = df["discount"] / 100
```

```
df['discount']
```

```
discount
0      0.0000
1      0.0000
2      1.3100
3      0.2700
4      0.0000
...      ...
27550   0.4980
27551   0.0750
27552   0.0000
27553   0.9900
27554   1.7547
```

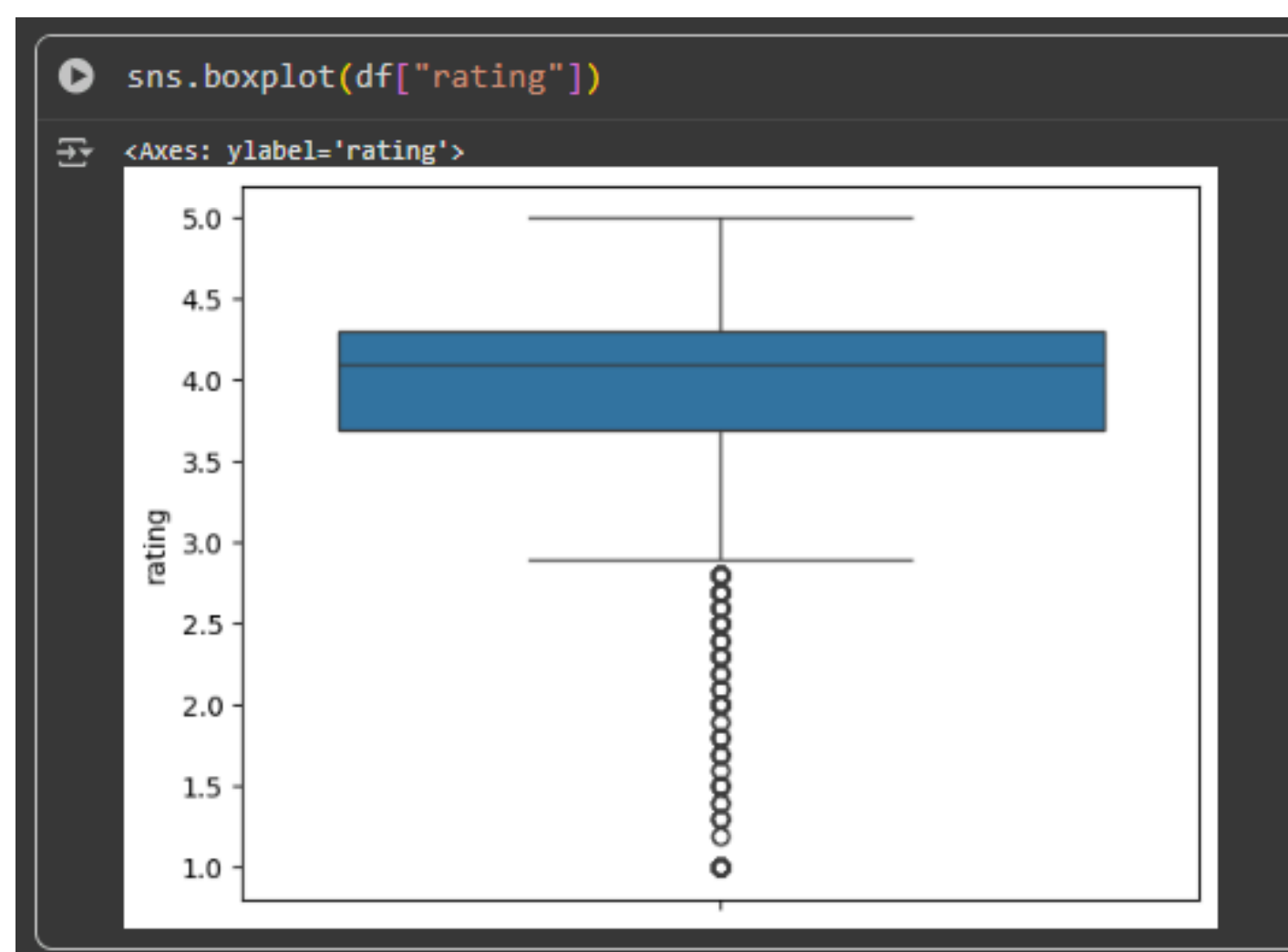
# Data Cleaning & Pre-Processing

- Find out the Missing Values from the Dataset.

```
df.isnull().sum()
```

	0
index	0
product	1
category	0
sub_category	0
brand	1
sale_price	6
market_price	0
type	0
rating	8636
description	115
products	1
discount	6

- Find out the outliers from the dataset according to the columns and fill them with the mean.



# Data Cleaning & Pre-Processing

- Find out the outliers from the dataset according to the columns and fill them with the mean.

```
df["rating"].fillna(df["rating"].mean(),inplace=True)
df["discount"].fillna(df["discount"].mean(),inplace=True)
df["products"].fillna("unkown",inplace=True)
df["description"].fillna("not_present",inplace=True)
df['brand'].fillna("unkown",inplace=True)
df["sale_price"].fillna(df["sale_price"].mean(),inplace=True)
```

Before cleaning

df.isnull().sum()	0
index	0
product	1
category	0
sub_category	0
brand	1
sale_price	6
market_price	0
type	0
rating	8636
description	115
products	1
discount	6

After cleaning

df.isnull().sum()	0
index	0
category	0
sub_category	0
brand	0
sale_price	0
market_price	0
type	0
rating	0
description	0
products	0
discount	0
dtype: int64	

# Data visulization

- Find out Top & least sold products.

```
df["products"].value_counts().sort_values(ascending=False).head()
```

products	count
Turmeric PowderArisina Pudi	26
Extra Virgin Olive Oil	15
Cow GheeTuppa	14
Soft Drink	12
Olive Oil Extra Virgin	12

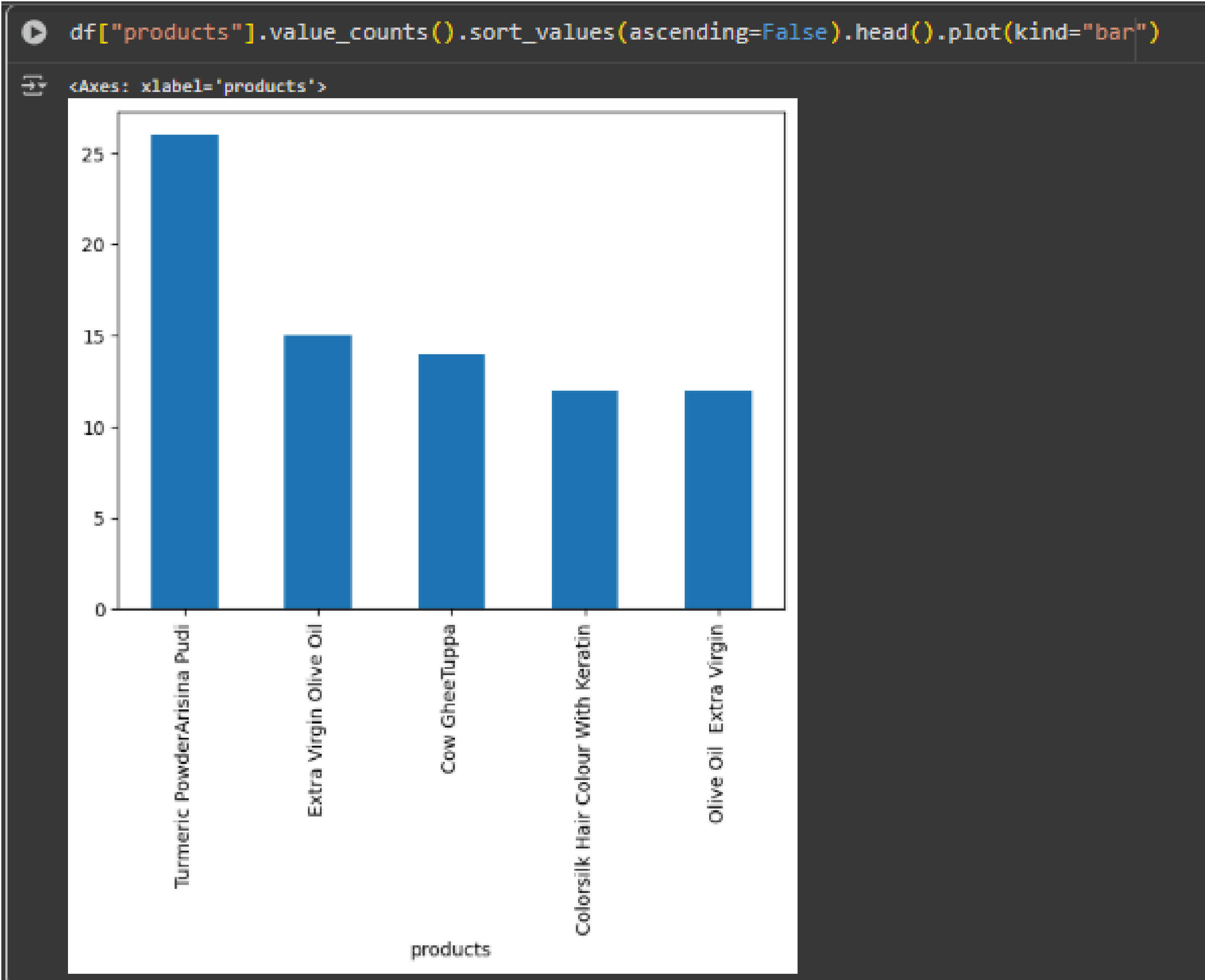
dtype: int64

```
df["products"].value_counts().sort_values(ascending=True).head()
```

products	count
W5 Perfume Spray For Women	1
Premium Care XXL 30 Diaper Pants	1
Dog Food Lamb Flavour	1
Hydra White Foam Cleanser	1
Safed Musli Root Powder Promotes Vigour & Vitality	1

dtype: int64

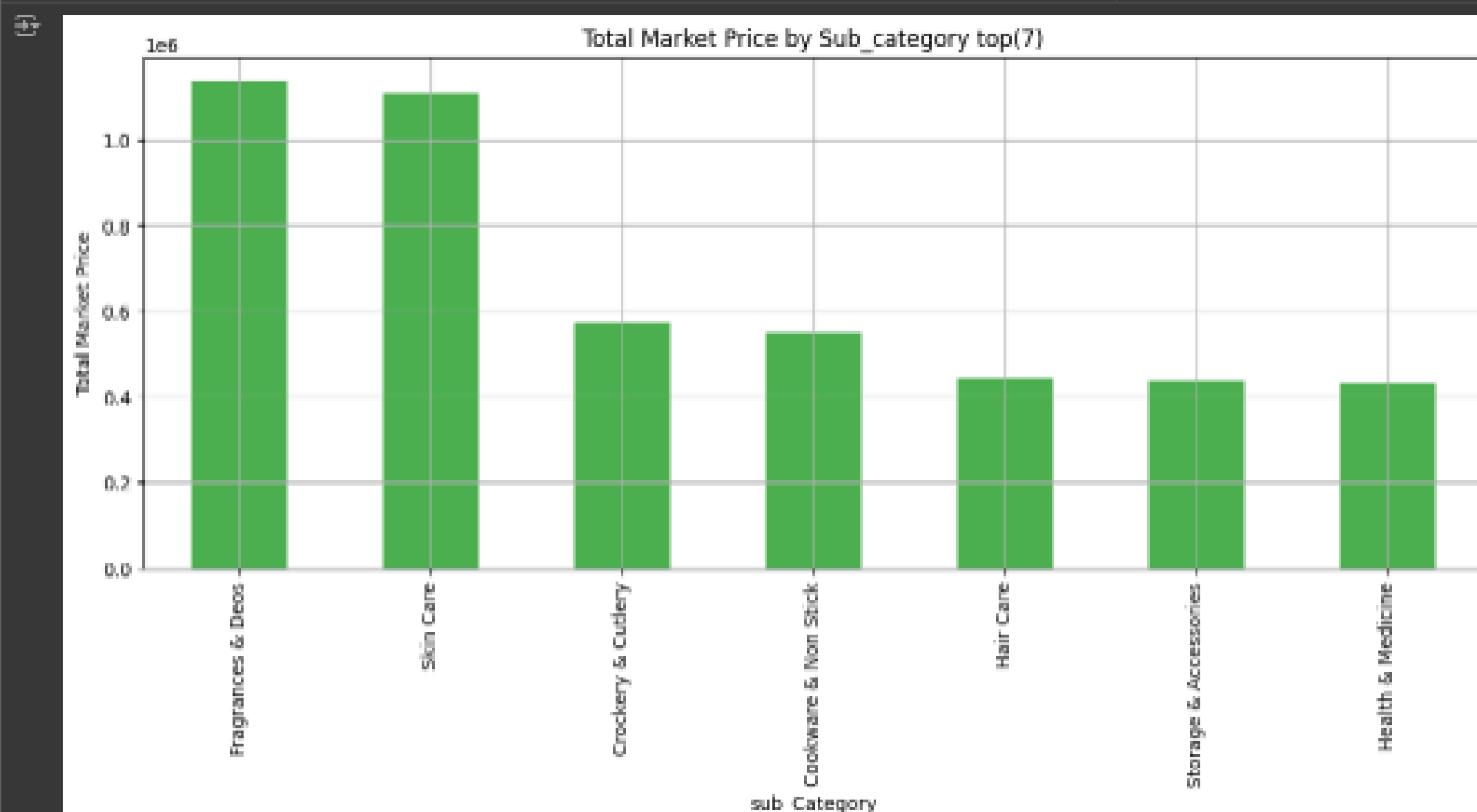
## Top (5) product sold



# Data visulization

- Top seven sub\_category by market price

```
plt.figure(figsize=(13, 5))
d = df.groupby("sub_category")["market_price"].sum().sort_values(ascending=False).head(7).plot(kind="bar",color = "#4CAF50")
plt.title("Total Market Price by Sub_category top(7)")
plt.xlabel("sub_Category")
plt.ylabel("Total Market Price")
plt.grid(True)
plt.show()
```



- Top seven sub\_category by Sale price

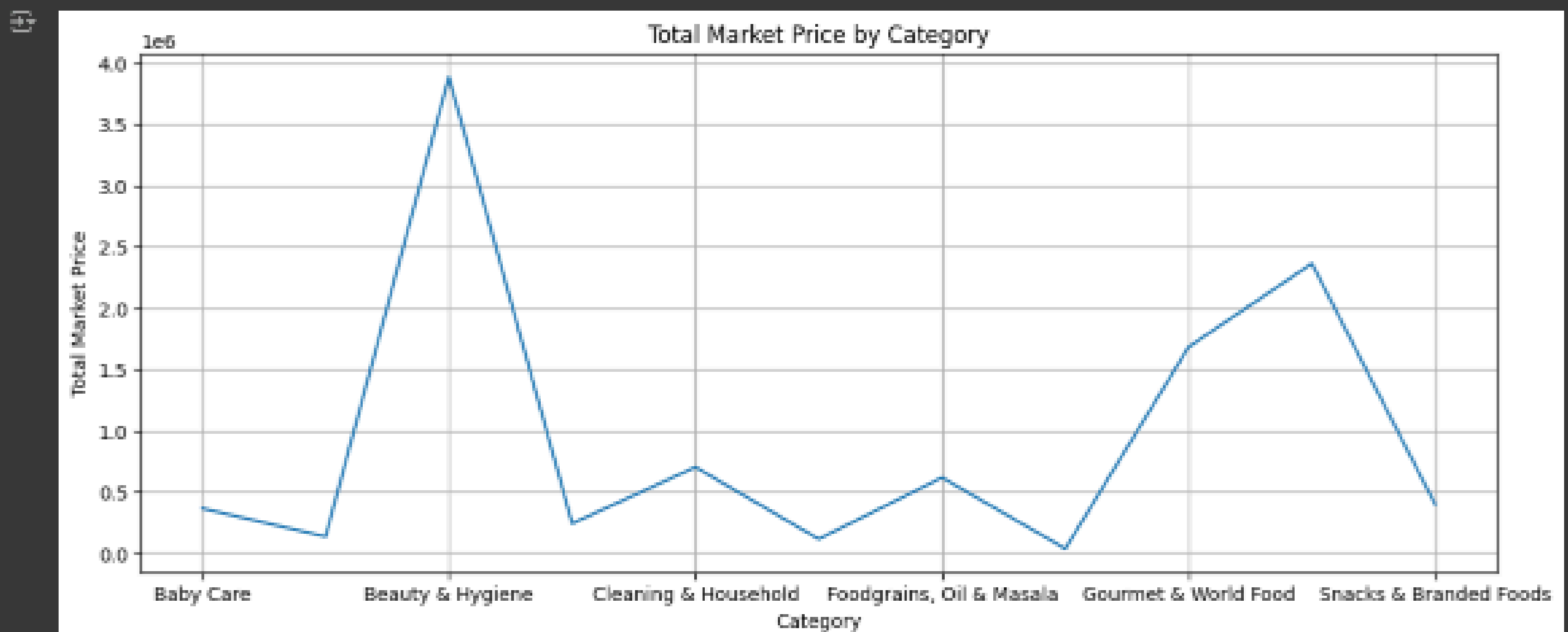
```
plt.figure(figsize=(13, 5))
d = df.groupby("sub_category")["sale_price"].sum().sort_values(ascending=False).head(7).plot(kind="bar",color = "hotpink")
plt.title("Total sale Price by Sub_category top(7)")
plt.xlabel("sub_Category")
plt.ylabel("Total Sale Price")
plt.grid(True)
plt.show()
```



# Data visualization

- Total market price by category

```
plt.figure(figsize=(13, 5))
d = df.groupby("category")["market_price"].sum().plot(kind="line")
plt.title("Total Market Price by Category")
plt.xlabel("Category")
plt.ylabel("Total Market Price")
plt.grid(True)
plt.show()
```



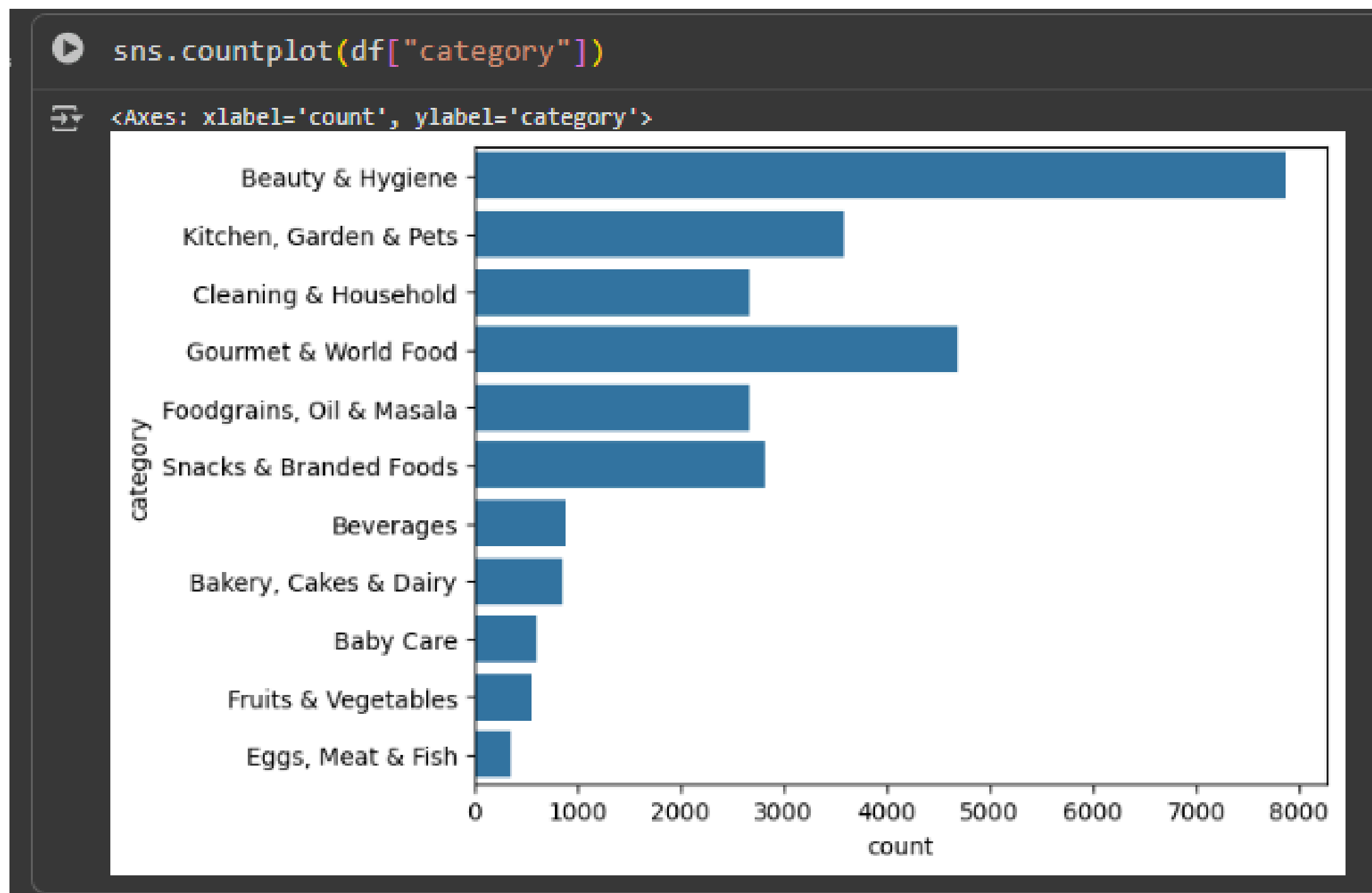
- Top seven brand by market price

```
plt.figure(figsize=(13, 5))
d = df.groupby("brand")["market_price"].sum().sort_values(ascending=False).head(7).plot(kind="bar", color = "#4CAF50")
plt.title("Total Market Price by Brand top(7)")
plt.xlabel("brand")
plt.ylabel("Total Market Price")
plt.grid(True)
plt.show()
```

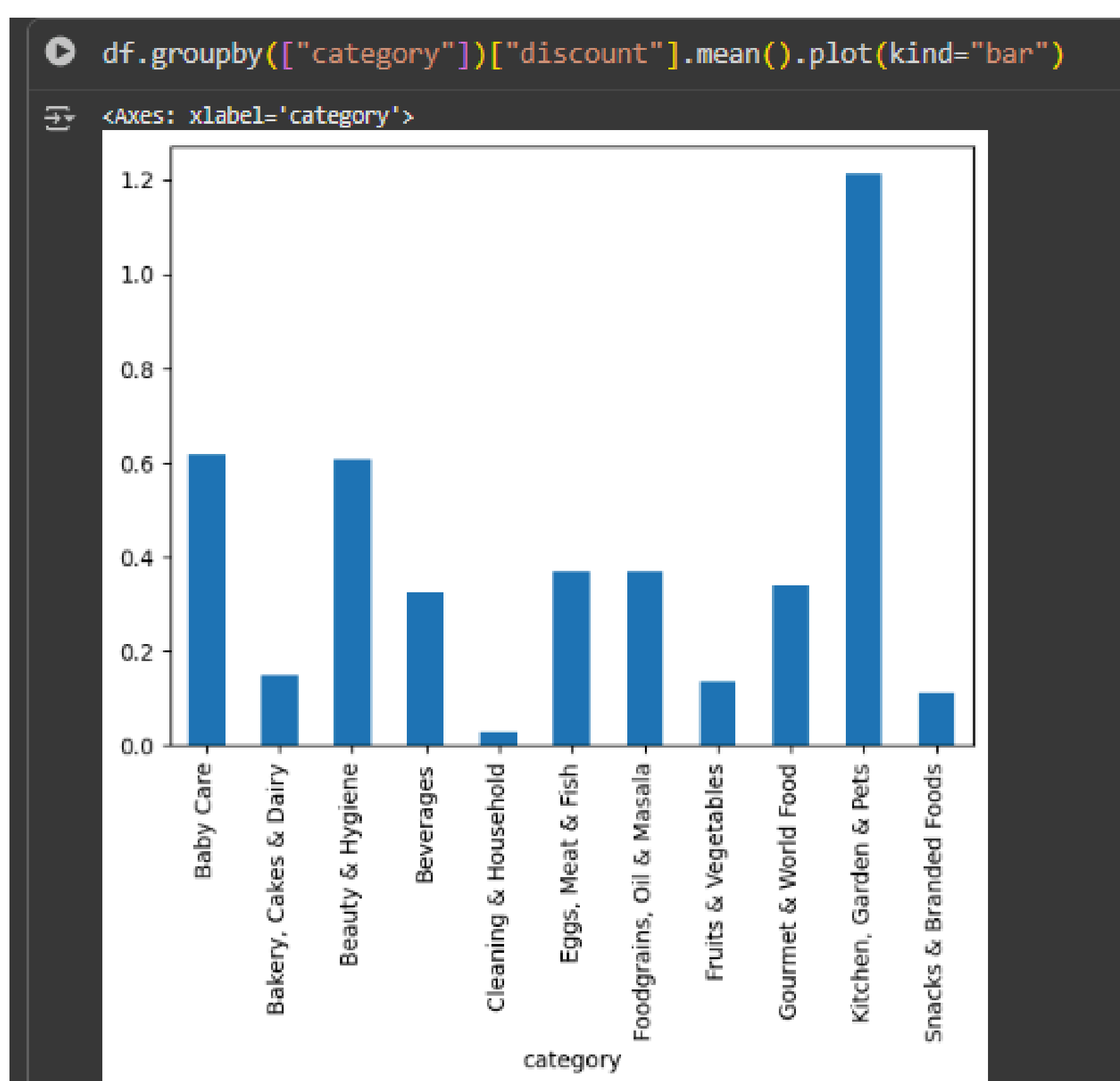


# Data visualization

- Count of category



- Category by discount



# Final report

The Big Basket E-commerce dataset provides a comprehensive overview of India's largest online grocery supermarket's product offerings, sales dynamics, and customer feedback. With ten key attributes encompassing product details, pricing information, brand categorization, and customer ratings, this dataset serves as a valuable resource for understanding the operational metrics and consumer preferences shaping the online grocery sector in India.

**The Big Basket e-commerce dataset** provides a robust foundation for comprehending India's online grocery market, offering invaluable insights into product demand, pricing approaches, customer feedback, and industry trends. By meticulously preparing the data, eliminating outliers, and conducting thorough exploratory analysis, this dataset empowers stakeholders to make well-informed decisions, streamline operations, and seize emerging opportunities in the fast-paced e-commerce landscape.

**In summary, the Big Basket e-commerce** dataset not only serves as a critical resource for immediate operational improvements but also lays the groundwork for long-term strategic planning. By harnessing the power of data analytics, stakeholders can navigate the complexities of the online grocery sector, adapt to changing market dynamics, and ultimately drive sustainable growth in this rapidly evolving landscape.

# ***Thanks for reading***



For coding part .....



Google Colab

[colab.google.com](https://colab.google.com)

